



# PREDIKCIJA U LINGVISTICI

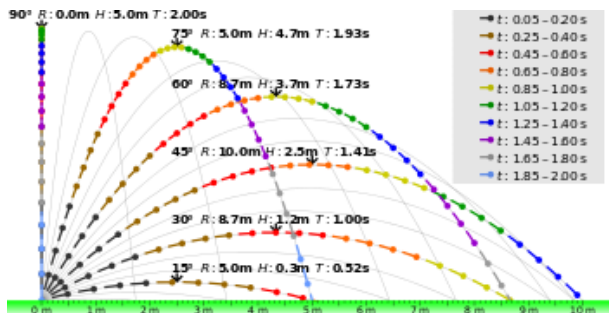
Tanja Samardžić

KorpusLab, PPI “Jezik i prostor”

21. 6. 2017.

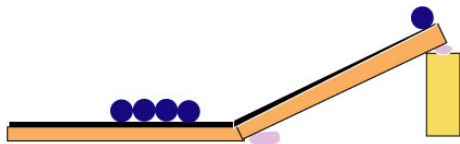
# Eksperiment i predikcija

## Putanja projektila u idealnim uslovima



# Opservacije i generalizacija

Put kuglice niz kosu ravan



Opservacije

$t$	$d$
1	0.5
1	1
2	2
3	?
4	7

Generalizacije  $d = 2^{t-1}$  ili  $d = \frac{1}{2}t^2$

# Predikcija u lingvistici

Šta se predviđa?

# Predikcija u lingvistici

Šta se predviđa?

- Gramatičnost rečenice  
npr. *\*Maj je procvetao ruže*

# Predikcija u lingvistici

Šta se predviđa?

- Gramatičnost rečenice  
npr. *\*Maj je procvetao ruže*
- Postojanje određenog elementa  
npr. “[...] we find verbs that lack anticausative counterparts in English (although their 'lexical semantics' would **predict** the existence of an anticausative variant)” (Alexiadou 2010)

# Predikcija u lingvistici

Šta se predviđa?

- Gramatičnost rečenice  
npr. *\*Maj je procvetao ruže*
- Postojanje određenog elementa  
npr. “[...] we find verbs that lack anticausative counterparts in English (although their 'lexical semantics' would **predict** the existence of an anticausative variant)” (Alexiadou 2010)
- Još primera?

# Statistička kontrarevolucija

## Pozitivizam 19. veka

Zakoni fonetski promena

## Strukturalizam, početak 20. veka

Potruga za objektivnim definicijama → deskripcija

## Kvantifikacija, 50-te godine 20. veka

Teorija informacija, Zipf

## Generativna gramatika 60-80-te godine 20. veka

Predikcija bez objektivnih definicija i merenja, dedukcija

## Razvijena statistika, 90-ih

Probabilistička predikcija, merenja, indukcija



# Suprotstavljanje teorije i podataka

## Kritika probablističkih modela (od strane teoretičira)

- Trivijalnost
- Zanemarivanje strukture
- Povezanost sa biheviorizmom: direktna veza stimulus-reakcija, bez unutrašnje reprezentacije

## Kritika teorijskog pristupa (od strane empirista)

- Parcijalnost
- Zanemarivanje jezičke raznovrsnosti

# Ka eksperimentu u lingvistici

U računarskoj lingvistici i veštačkoj inteligenciji promena od 90-ih  
(Abney 2011, Russel and Norvig 2002)

- Predlaganja novih teorija → razrada postojećih
- Intuicija → teoreme ili solidni empirijski dokazi
- ?? → replikativnost

## Primer: leksički kauzativi

Bez alternacije  
(samo neprelazni)

Sa alternacijom  
(i neprelazni i prelazni)

---

*rasti*  
*(pro)cvetati*  
*(o)stariti*

*smanji(va)ti (se)*  
*(na)smejati (se)*  
*(pod)mladiti (se)*

# Pretpostavka Alexiadou (2010)

## Klasa I

u osnovi neprelazni

interno prouzrokovani i nespecifikovani

---

*(po)beleti (se)*

*(za)obliti (se)*

*(pro)širiti (se)*

*(o)sušiti (se)*

*otvoriti/otvarati (se)*

*(pro)cvetati*

## Klasa II

u osnovi prelazni

eksterno prouzrokovani

---

*(po)cepati (se)*

*umanji(va)ti (se)*

*uništiti/uništavati (se)*

*(za)paliti (se)*

*(sa)kupiti (se)*

*(po)tonuti (se)*

*(iz)lečiti (se)*

# Pitanje: Koji tip glagola u kojim jezicima alternira?

## Generalizacija

Jezici	Grupa A	Grupa B
Tipovi		
Specifikovan (bilo interno bilo eksterno prouzrokovan)	NE	DA
Nespecifikovan	DA	DA

Jezici: gruzijski, mongolski, ruski, poljski, japanski, "Salish", engleski, grčki, hindi, turski

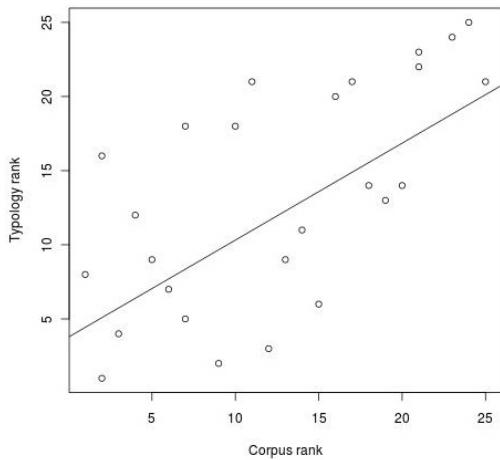
# Metodološki propusti

- Kriterijum za odabir uzorka glagola
- Kriterijum za odabir uzorka jezika
- Ne postoji provera generalizacije

# Probabilistički pristup leksičkim kauzativima

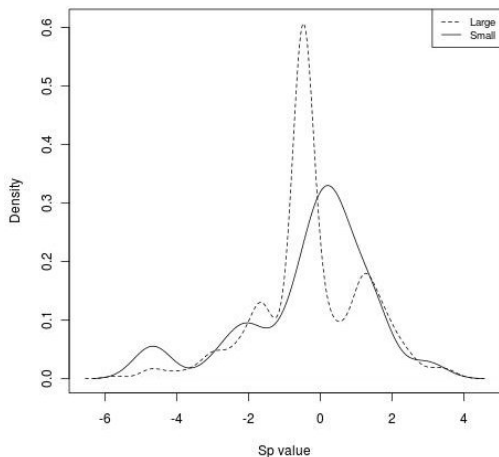
- Uzorak glagola:
  - tipološki pregled (N=26)
  - jednojezični indeks (N=354)
- Uzorak jezika
  - tipološki pregled (N=21)
  - korpus jednog jezika
- Merenje i predikcija
  - Korpusna mera tranzitivnosti kao indikator tipa prouzrokovanosti ( $Sp$ -value)
  - Tipološka mera morfološke markiranosti kao indikator tipa prouzrokovanosti
  - Koeficijent korelacije između dve mere
  - Distribucija mere tranzitivnosti u glagolima jednog jezika

# Korelacija između dve mere





# Distribucija korpusne mere tranzitivnosti



## Koji tip glagola u kojim jezicima alternira?

Što je  $Sp$ -value bliža nuli, to je veći broj jezika u kojima glagol alternira.

Redefinisanje pitanja: da li ima smisla predviđati postojanje alternacije za bilo koji konkretan glagol u bilo kom konkretnom jeziku?

Steven Abney. 2011. “Data-intensive experimental linguistics”.  
*Linguistic Issues in Language Technology — LiLT*, 6(2), pp.1–30.  
<http://elanguage.net/journals/lilt/article/view/2578>

Stuart Russell and Peter Norvig, 2002, “Artificial Intelligence: A Modern Approach”.